Housing Market 101

By Bhupesh Kumar

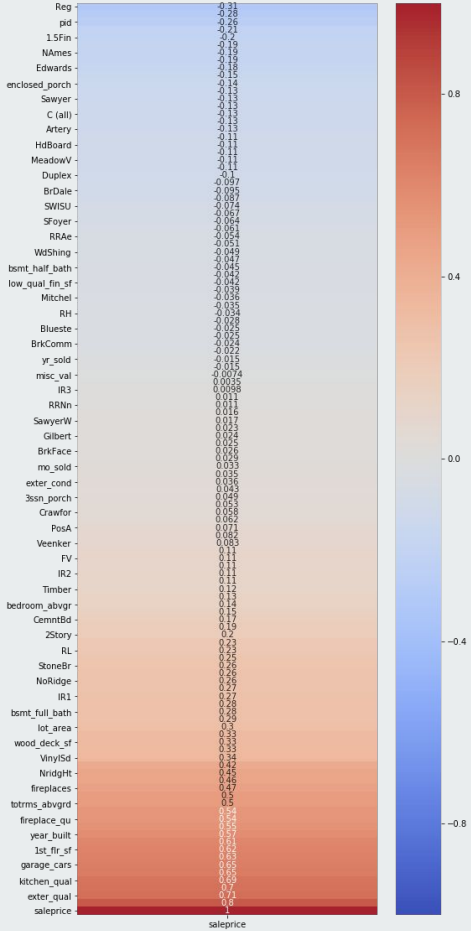
Assumption

- Location Location Location
- Number of Bedroom
- Size of the house
- Number of Bathroom in the house
- Neighborhood
- Year built
- Year remodeled

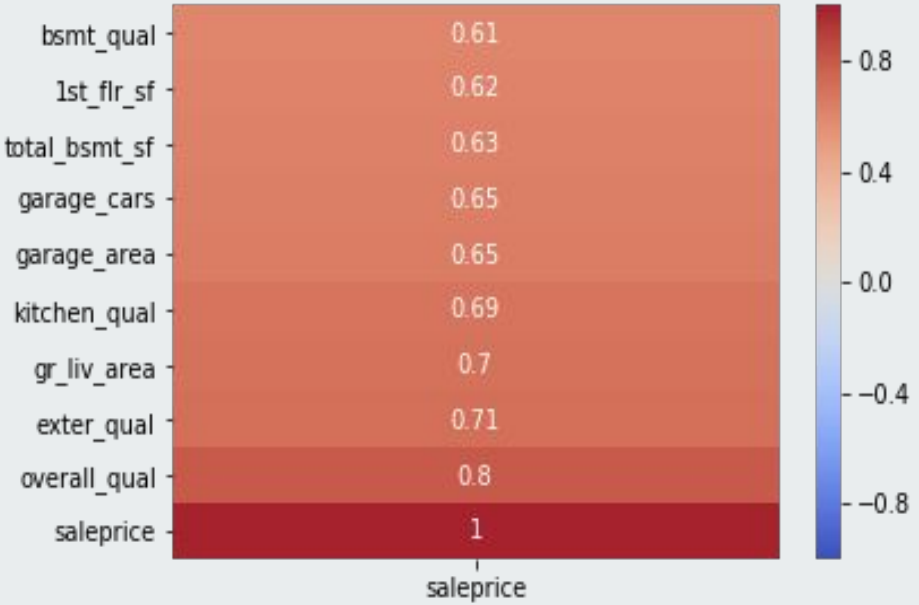
AMES Housing data

Data Description :

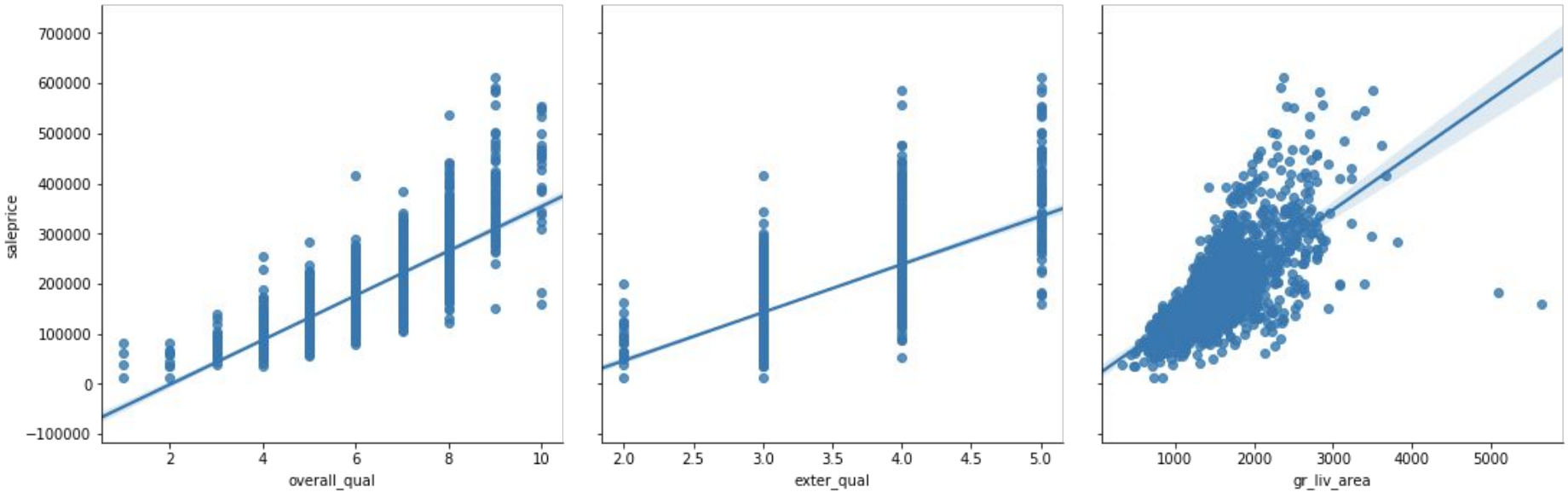
- Ames , IA
- Data Collection from 2006 -2010
- 2051 Rows and 82 Columns of data ~ over 150,00 data entries
- Mix of data types : Nominal, Ordinal, Discrete and continuous

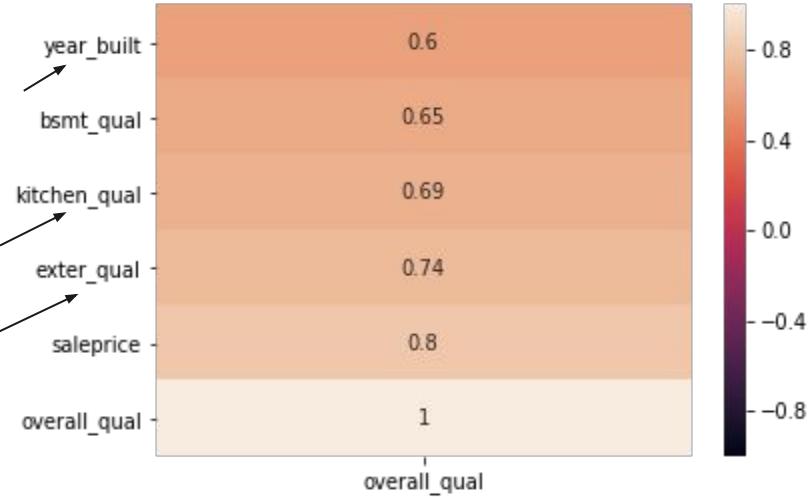


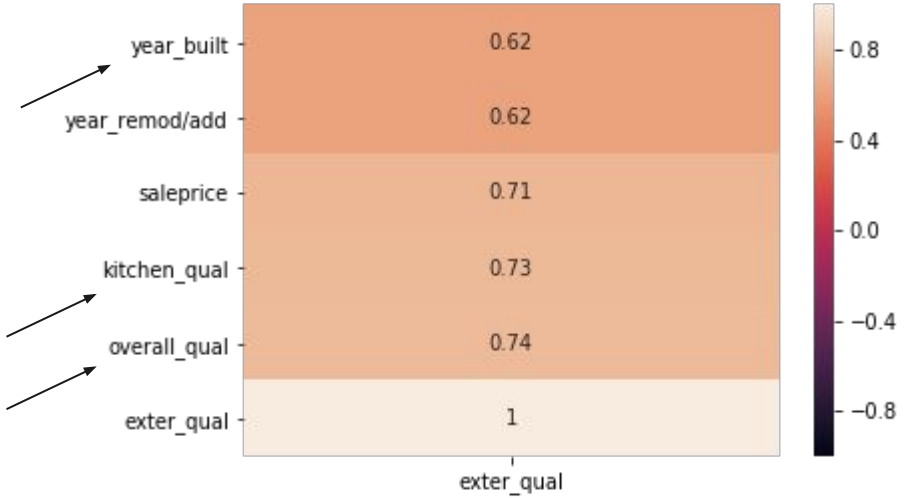
Heat Map showing the Correlation of all columns against Sale Price

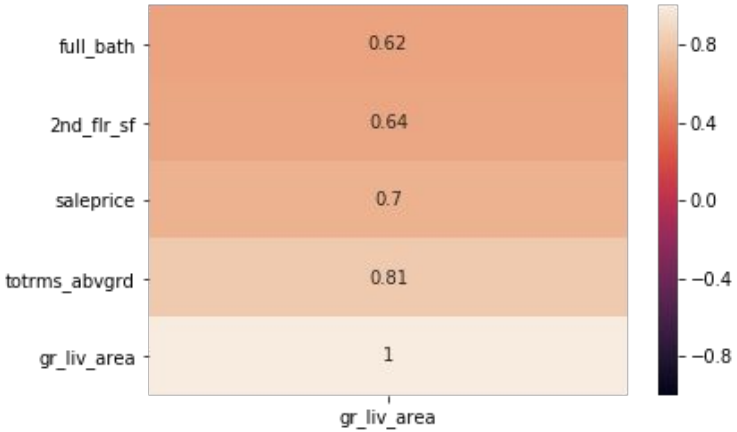


Strong Features correlate to Sale Price

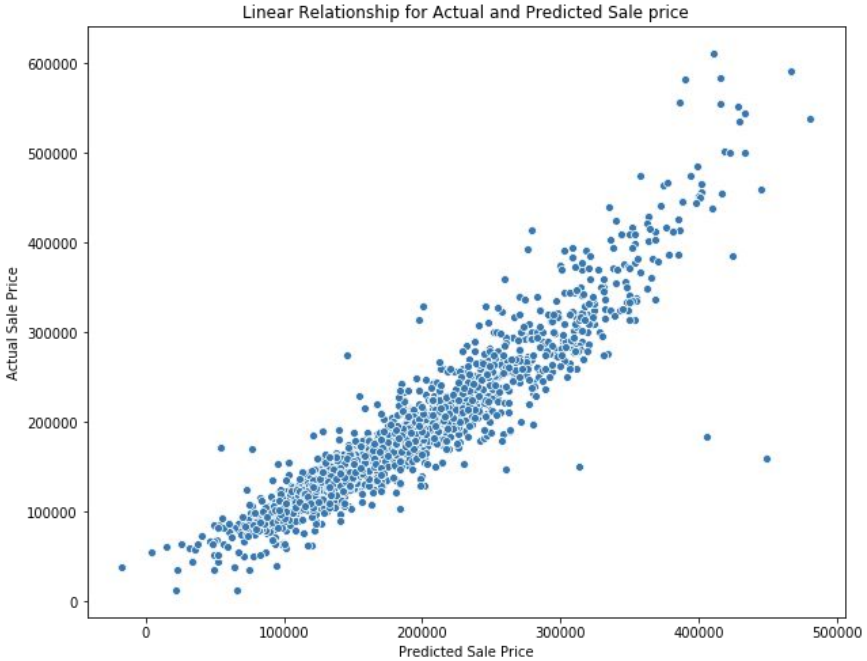



year_built
bsmt_qual
kitchen_qual
exter_qual
saleprice
overall_qual
0.6
0.65
0.69
0.74
0.8
1
overall_qual
0.8
0.4
0.0
−0.4
−0.8


year_built
year_remod/add
saleprice
kitchen_qual
overall_qual
exter_qual
0.62
0.62
0.71
0.73
0.74
1
exter_qual
0.8
0.4
0.0
−0.4
−0.8

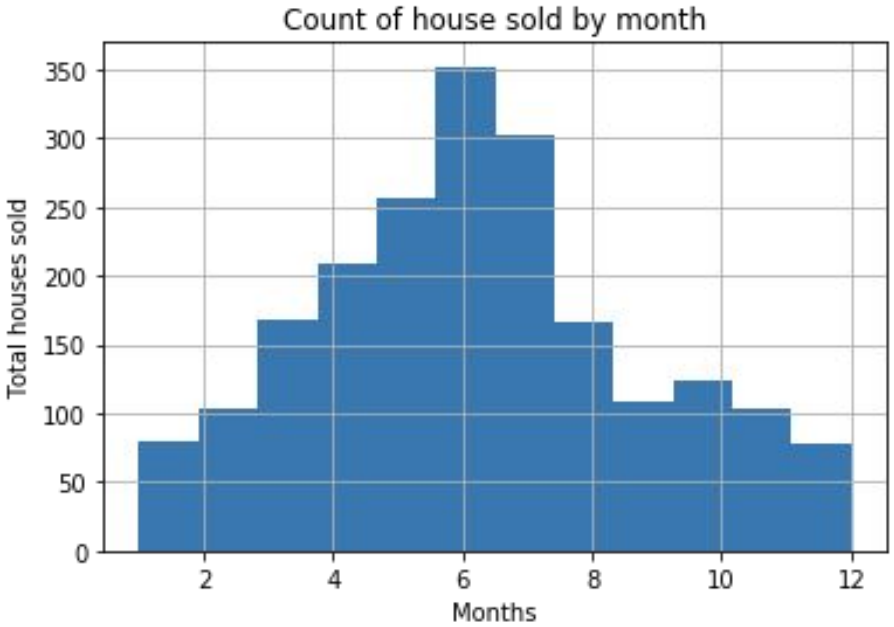

full_bath
2nd_flr_sf
saleprice
totrms_abvgrd
gr_liv_area
0.62
0.64
0.7
0.81
1
gr_liv_area
0.8
0.4
0.0
−0.4
−0.8

Conclusion



Using LassoCV Model, I was able to get R2 score of 88 %.

Food for thought



54% of houses are sold between months of April, May, June, and July

April - 208
May - 257
June -352
July -303

Thank you !!